Alexandria Engineering Journal (2022) 61, 7659–7668

HOSTED BY

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

Research on user generated content in Q&A system and online comments based on text mining

Yahui Chen, Dongsheng Liu, Yanni Liu *, Yiming Zheng, Bing Wang, Yi Zhou

Zhejiang Gongshang University, Hangzhou, China

Received 8 December 2021; revised 4 January 2022; accepted 5 January 2022
Available online 14 January 2022

KEYWORDS

Q&A system;
User generated content;
Content classification;
Cluster analysis

Abstract For information asymmetry in e-commerce platforms, question answering system (Q&A) and online comments are used to build trust relationship. Although online reviews play a significant role in reducing information asymmetry, information overload and distortion are also not negligible. In the study, text mining was used to extract various questions from Q&A system and online comments to obtain the content that consumers are most concerned about in online shopping, and the similarities and differences of the two mining results are compared. The results show that Q&A system has a great influence on users' decision making. The Q&A system is not only complementary to online comments, but also can provide validation for the information of online comments.

1. Introduction

With the rapid development of Internet in recent years, online shopping has become the main choice for consumers. However, the phenomenon that the products are inconsistent with the description provided by merchants is very common, which makes consumers gradually lose trust in online shopping and hinders the development of e-commerce to a large extent [1,2,3]. Then the online comments, question answering system (Q&A) and other services are provided on the e-commerce platform. They can provide consumers with the information of products, which can directly affect consumers' shopping decision [4]. The combination of information provision and product recommendation in online reviews has played a significant role in relieve information asymmetry [5], but the surge in the number of reviews and the inappropriate interference behavior of merchants have also resulted in information overload and distortion in online reviews [6]. By contrast, the product information obtained through the Q&A system has powerful pertinence and save time and effort. However, there are less studies on the difference between the Q&A system and online comments in the existing literature. For example, what are the characteristics and specific attributes of the information content in the online comments and Q&A system? What's the difference in both of them?

In the study, the laptop products in Amazon are chosen as the experimental object. Firstly, the text data of Q&A system is classified and clustered to mine the characteristics of the information content and determine service category. In order to further confirm the unique attributes in different types of user generated content, the online comment of the laptop commod-

* Corresponding author.
E-mail addresses: yahuic@zjgsu.edu.cn (Y. Chen), lds1118@zjgsu.edu.cn (D. Liu), yanniliu@zjgsu.edu.cn (Y. Liu), zheng20232@163.com (Y. Zheng), Iceking0706@zjgsu.edu.cn (B. Wang), zyhello@mail.zjsu.edu.cn (Y. Zhou).

Peer review under responsibility of Faculty of Engineering, Alexandria University.

https://doi.org/10.1016/j.aej.2022.01.020

ity is clustered and analyzed. Moreover, this paper summarized the characteristics of user generated content compared with the mining results of the Q&A system and online comments.

This paper contributes to analyze the characteristics of user generated content in Q&A module and online comments on e-commerce platforms, which can help consumers make comprehensive use of different user-generated content platforms to quickly find the requirement information and make shopping decisions. Secondly, it is helpful for the e-commerce platform to make improvement in the management of user-generated content. The questions and answers concerned by consumers can be further classified and integrated to save time and effort, and improve the shopping experience. Finally, it is helpful for the e-commerce platforms to clarify the real needs of users, and improve product design, product quality and user experience.

2. Related works

2.1. Q&A system of e-commerce

The Q&A system on the e-commerce platform is a new type of online Q&A community form, which uses the community wisdom to generate content and conduct knowledge exchange by asking and answering questions. In the study of the online Q&A community, Bickart and Schindler demonstrated that online forums are the important source of user generated content [7]. In addition, the study found that user generated content had a significant impact on stock performance in finance [8].

Considering the advantages of online Q&A communities, e-commerce platforms have launched similar services, such as the Q&A system of amazon, which allows consumers to ask questions or give answers on the platform. Participants can ask and answer questions about products. The research of Q&A system mainly involves the correctness of answering information and user's behavior. Previous studies have attempted to distinguish the quality level of answering, and found that there were some differences in the quality of answers provided by different websites based on the problems and theories in the field of information retrieval [9]. For example, Choi and Shah identified the motivations to ask a question in online Q&A environments [10], which contributed to develop question-answering processes and gain insights. Yu et al. studied the impact of social and technological mechanisms such as user incentive reputation system on knowledge contribution [11]. From the perspective of motivations, Fang and Zhang studied the continuous participation behavior in online social Q&A communities based on theory of planned behavior [12]. Most existing researches focus on the behavior and quantity of knowledge contribution behavior of users, but lack the researches on the quality of knowledge contribution.

2.2. Online comments of e-commerce

The existing research on online reviews can be divided into two directions. On the one hand, the research on generated online reviews mainly analyzes the characteristics of online reviews from the perspective of usefulness, authenticity and deviation of content. At the same time, the impact of online reviews on product sales and performance is analyzed [13]. On the other hand, the existing research doesn't explain how to make users have a higher intention to publish comments, and analyzes the factors and motivations to publish comments.

Online reviews have been taken as the most important way to impact the consumers' information acquisition and decision process [14]. When browsing on e-commerce platforms, consumers usually pay attention to the reviews provided by people who have bought the products, which are regarded as more valuable and credible information [15]. The comment content allows potential consumers to gain insight into the product and make decisions [16]. Mosteller and Mathwick identified online reviews as the unique form of word-of-mouth marketing, and found that the impact degree of online reviews on consumers' shopping decisions reached 20 to 50 percent [17]. In addition, different from the effect of word-of-mouth, online reviews can break through the limitations of community [18].

Although the large number of comments provide more information to consumers, consumers may spend more time and efforts in searching for useful information [19,20]. Such information overload will lead to the increase of consumers' cognitive cost, which may lead to the intention reduction of consumers' reading comments. Ghose and Ipeirotis used the random forest algorithm to solve the problem of sorting comments [21]. Qiu et al. used the syntactic principles to find syntactic similarity on feature words and locate the features related with sentiment words in document reviews [22]. In order to extract some hidden features from online comments and provide more accurate and detailed text mining results, Lazhar proposed a classification-based approach to extract implicit opinions [23].

2.3. Text mining on comments

In the era of big data, the increase of information is particularly obvious in e-commerce platforms. Even the products with low sales have thousands of comments. The problem of information overload in online reviews not only affects the usage of reviews, but also reduces the value of reviews. While user generated content is vital in purchase decisions in e-commerce, refining summary of user generated contents is necessary [24]. To this end, text mining method is used to find more valuable information from online comments. As a method of text analysis, text mining is usually combined with statistics and machine learning to mine useful knowledge from text [25]. The key steps and main contents of text mining include text preprocessing, classification, clustering, and visualization of results [26,27], which can help people find interesting contents from massive texts [28].

In the study, we will mine the information content characteristics of e-commerce Q&A system through text preprocessing, classification, clustering and visualization. Text preprocessing, as the key step in text mining, is usually word segmentation, word removal and feature extraction [29]. In addition, in view of the comments of short text, clustering methods mainly include the extension and selection of features. For the words' structure and the inaccuracy in short text, the existing research focuses on characteristics of extension. Through semantic dictionary or opening searching engines on the sparse characteristics of expansion and transformation, the correlation among texts can be enhanced, which makes the

clustering results more accurate. For the low efficiency of short text clustering in word extension, the text representation method based on feature keyword extension is proposed to improve the accuracy and efficiency of short text clustering [30]. In the specific context, the chosen clustering algorithm should be suitable for the characteristics of the clustered text.

3. The proposed method

3.1. Data collection

This paper selects laptop products from the global Amazon platform as the research data, which provides consumers with Q&A system and online reviews. The crawling data of the Q&A module mainly includes questions, answering contents, number of answers, questioners' ID, respondents' ID, question time, answering time. The Q&A module is shown in Fig. 1. This paper crawled 90,886 questions and answers from 1,822 laptop products in Amazon, covering 30 computer brands. At the same time, this paper crawls online reviews of these laptops, including user name, membership level, rating star, comment content, comment time, the number of likes, the number of comments, and page title. Due to the large amount of online review, this paper only crawls the first 100 reviews for each laptop product. Eventually 147,462 online review texts are crawled from 1,547 products with online reviews.

3.2. Data preprocessing

Since the data crawled in this paper is unstructured, we firstly preprocess the data by deleting duplication, empty content in the text. Then text is preprocessed by word segmentation and removal of stopping words. The specific process is shown in Fig. 2.

When word segmentation and the removal of stop words are finished, the clustered text is converted to word frequency matrix. And we use word cloud to visualize the clustering results. According to the word cloud, some noise words can be found and added to the stop words list. Through the clustering experiments, all the noise words were obtained. Meanwhile, the words with the highest frequency were traced back, and the words matched were found based on the word clouds. Then the new words were added into the original word segmentation dictionary.

3.3. Content classification based on LSTM

Given that the large amount of Q&A text usually involves the characteristic words in other categories, which is difficult to be extracted by clustering method. In order to remove the category of commodity comparison from the question texts, the binary classification method is used to classify the question texts of e-commerce Q&A system.

Long-Short Term Memory (LSTM), as a kind of Recurrent Neural Networks (RNN), not only improves the deficiency of

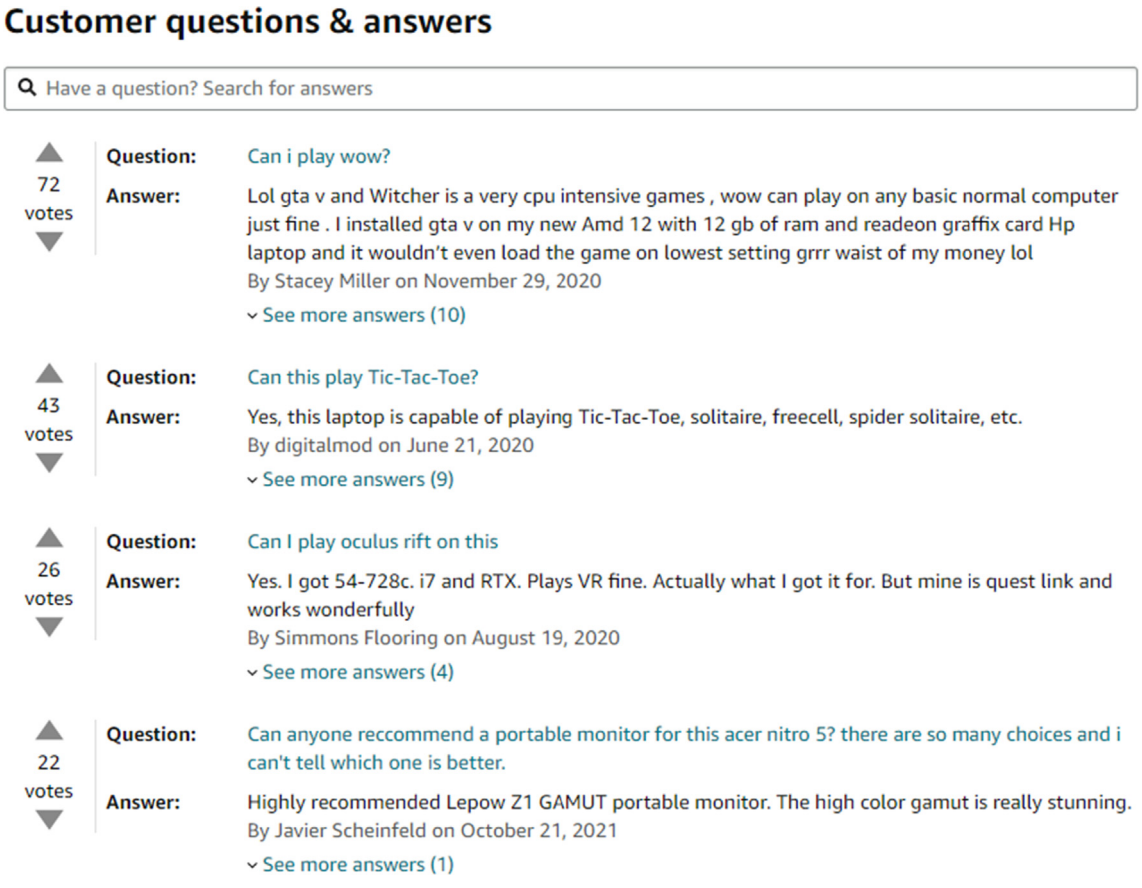


Fig. 1 Q&A system module.

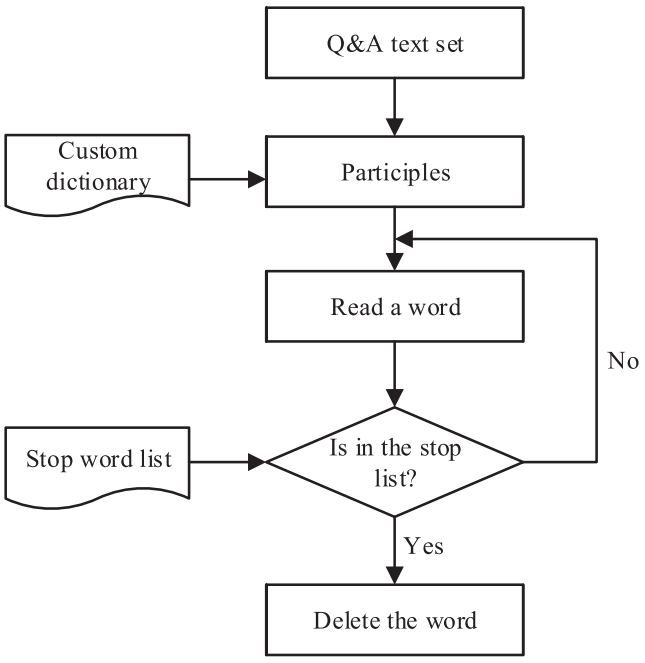


Fig. 2 Text preprocessing of Q&A system.

RNN in long sequence data processing [31], but also has good memory ability and avoids the disappearance of gradient. For the classification of short text, LSTM can use context feature information to automatically select features for classification after remembering the feature information [32,33]. In order to determine whether the information is useful, a cell structure is added to improve LSTM algorithm. As is shown in Fig. 3.

The previous research shows that LSTM algorithm was better than traditional SVM and RNN algorithms in text classification [34,35]. In the study, LSTM is selected as the classification model. And the trained model is used to process the original data so that 6,224 pieces of data in the category of commodity comparison are separated, while the remaining 84,662 pieces of question content text in this category are not compared. The LSTM architecture used in this experiment is as follows:

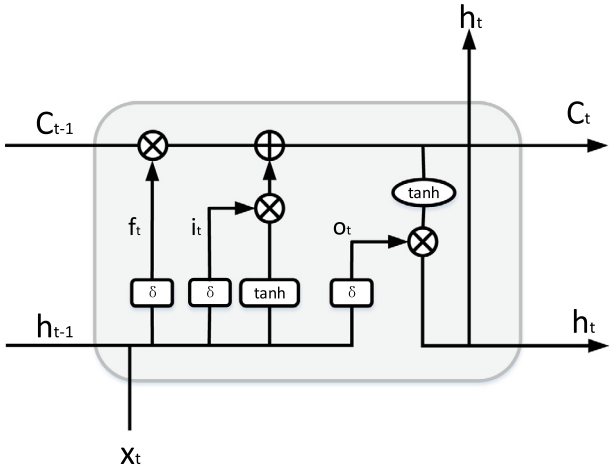


Fig. 3 Schematic diagram of long-short term memory.

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (1)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (2)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (4)$$

$$h_t = o_t \tanh(c_t) \quad (5)$$

Where f_t, i_t, c_t, and o_t correspond to the forgetting gate, input gate, the activation vector of cell state, and output gate. These vectors are the same in size as the hidden output vector h_t, and the hyperbolic tangent function is represented by tanh. The weight matrix is represented by W. W_{xo}, W_{xf} stand for the weight matrix of the output gate and forgetting gate. σ represents sigmoid activation function.

3.4. Content clustering based on the improved K-means algorithm

The initial clustering center in the traditional K-means algorithm is randomly selected, but the selection of the initial clustering center has a great impact on the clustering result [36,37]. Therefore, K points far away from each other are selected as the initial clustering center to improve the algorithm in the paper. The specific algorithm steps are as follows:

Input: Initial sample set $X = \{x_1, x_2, ..., x_n\}$ and number of class clusters k.

Output: K class clusters conform to the convergence of error square sum criterion function.

Step 1: Randomly select a sample from the sample set as the initial clustering center c_1.

Step 2: For each point x_i in the sample set, calculate the distance to the nearest clustering center $D(x)$.

Step 3: Select a new sample as the new clustering center. The rule of selection: the larger the $D(x)$ is, the higher the probability of being selected as the clustering center.

Step 4: Repeat the above two steps until k clustering centers are selected.

Step 5: For all samples in the sample set, calculate the Euclidean distance from each sample to cluster center and classify it into the cluster where the cluster center with the minimum distance is located. For each cluster c_i, the cluster center $c_i = \frac{1}{|c_i|}\sum_{x \in c_i} x$ is updated, and $|c_i|$ is the number of such samples. Until the center of the cluster remains unchanged, the sum of the squares of errors will converge.

Step 3 is important to the improve K-means algorithm and the computing method on the probability of the selected point $D(x)$ is as follows. Add the distance between each point in the data set and the nearest clustering center point, and the sum is denoted by $sum(D(x))$. A random value $Random$ between 0 and $sum(D(x))$ is selected, and $Random = Random - D(x)$ is computed until $Random \leqslant 0$. Then this point is regarded as the next center of the cluster.

In this paper, the improved and original algorithm are used to cluster the test data and the sum of error squares is used as the evaluation criteria. The sum of square errors $SSE = \sum_{i=1}^{k}\sum_{x \in c_i} dist(c_i, x)^2$ is used as the evaluation criterion of the algorithm. Moreover, dist represents the distance from each point to the center point of the cluster to which it belongs,

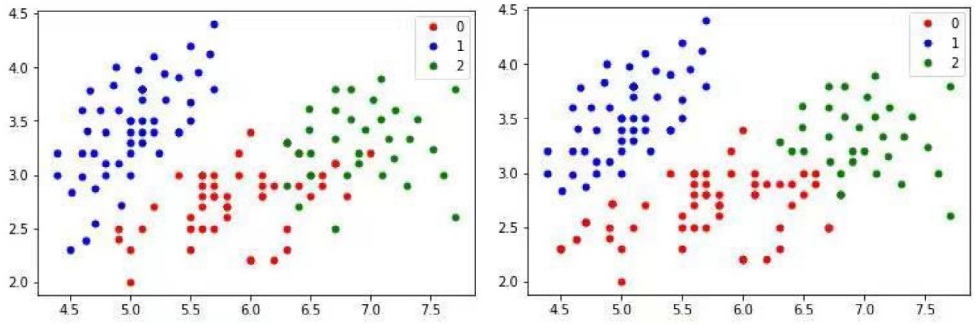
Table 1 Comparison of SSE values in two algorithms.

Iterations	K - means algorithm	Improved K-means algorithm
2	768.52	238.59
3	736.68	232.67
4	584.16	232.67
6	247.48	232.67
8	232.67	232.67
10	232.67	232.67

and SSE value is the sum of the squares of the distance from all sample points to the center point of the cluster to which it belongs. In the same number of iterations, the smaller the SSE value is, the better the algorithm is and the smaller the information loss is.

As is shown in Table 1, traditional K-means algorithm converges to the minimum at the 8th iteration, and the SSE value is 232.67. However, the improved K-means algorithm converges to the minimum at the 3rd iteration. It shows that the clustering effect of the improved K-means algorithm is better than that of the k-means algorithm in Fig. 4, which can more clearly divide the original data points into three clusters according to the distance similarity. Thus, we found that the efficiency of the improved K-means algorithm has been improved, which can rapidly and stably converge to the minimum with fewer iterations, and the clustering result is better.

The improved K-means algorithm of this paper classifies the comments to the behavior after clustering. Since the algorithm belongs to unsupervised learning, it can't be known the final number of clusters before clustering. But according to the pretreatment of the early work, it can be determined the final number of at least five clustering results. Therefore, by setting the parameter n_clusters = range (5, 20, 1), the text data in the question data of Q&A system was clustered. The final clustering number K is determined by calculating the contour coefficient. The higher the contour coefficient is, the higher accuracy the clustering result is. In the process of clustering, the text should be firstly segmented and a word segmentation dictionary should be constructed by adding laptops' noun in color, model, hardware, software and other terms. Then, the tfi-df value of each word is calculated and the text to be clustered is converted to the word frequency matrix. In this paper, Matplotlib was used to visualize the result of the first cluster-



(a) K-means algorithm (b) Improved K-means algorithm

Fig. 4 Comparison of clustering effect.

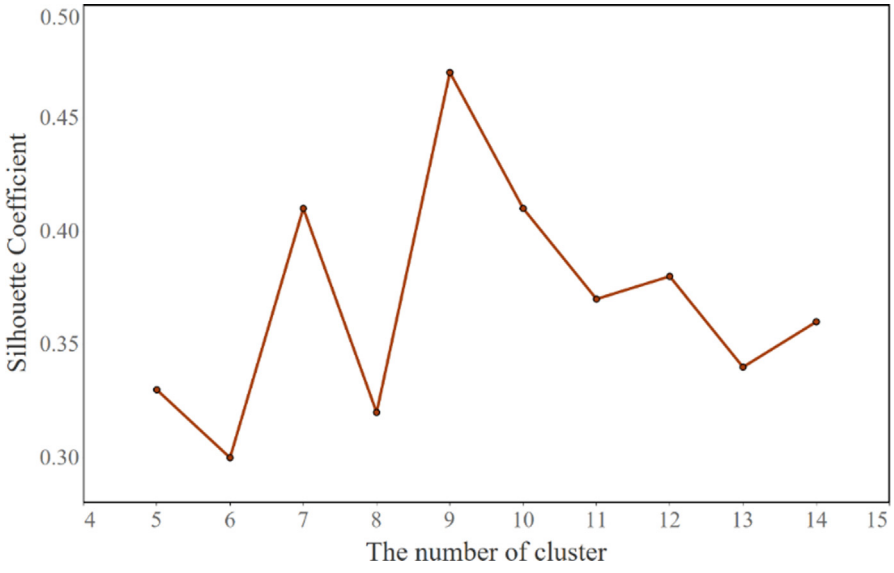


Fig. 5 Silhouette coefficient of different number of cluster in Q&A system.

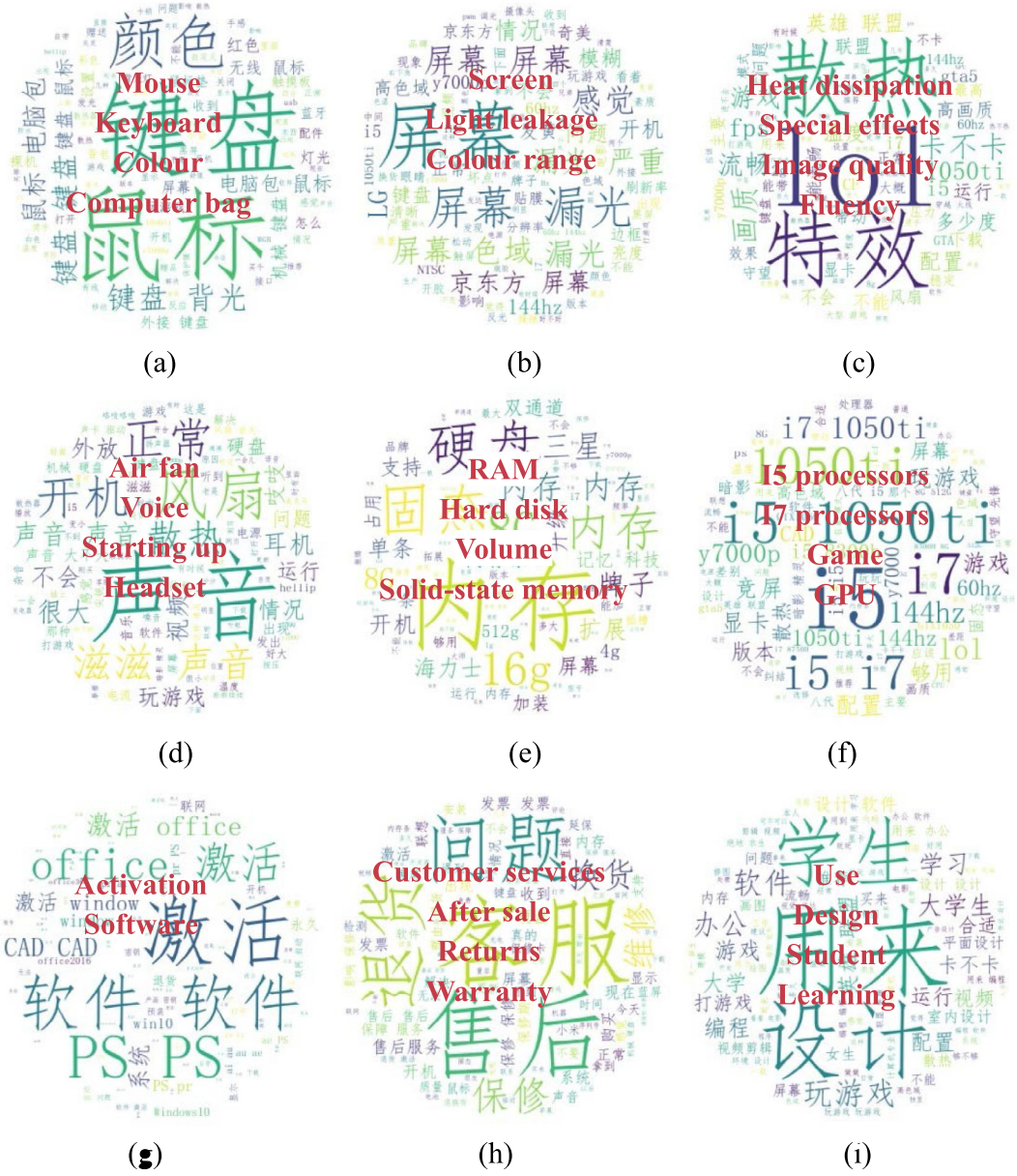


Fig. 6 Cloud diagram of clustering words.

ing, and it was observed that there were many noisy words in the generated word cloud map. In order to make the final clustering effect more accurate, these noisy words were added into the stop words table to remove. Finally, the clustering text is analyzed to obtain the information characteristics and types of user-generated content.

4. Experimental results and analysis

4.1. Features of user generated content in Q&A system

In order to remove the question content of commodity comparison from all the question and answer content, this paper input all the text content of 90,886 question and answering into the trained LSTM model. The 6,224 question and answering content of commodity comparison is separated. And the remaining 84,662 question and answer content was obtained. In order to improve the accuracy of the clustering result, this paper screened 84,662 Q&A system text, and finally got 81,329 Q&A system text data. In this paper, the optimal clustering number is determined according to the silhouette coefficient. The changing diagram of silhouette coefficient is constructed with different value of cluster, as shown in Fig. 5.

As it can be seen from Fig. 5, when the number of cluster K is 9, the silhouette coefficient reaches the peak, and the clustering effect is the best. Combined with the text data analysis of the Q&A system, this paper constructed the clustering word cloud diagram of the Q&A system for the selected commodity, as shown in Fig. 6.

Through the clustering results, the feature words of user-generated content in the five category Q&A system is shown in Table 2.

Table 2 Feature word clustering in Q&A system.

Clustering categories	Main features
Gifts	mouse, keyboard, computer bag, mouse pad, colour
Quality	screen, light leakage, blur, bad point, clear, color gamut, resolution, black screen, special effects, fluency, image quality, cooling, crash, sound, boot, fan, 1050ti, i5 processors, i7 processors, 8G, 144hz, graphics, hard disk, memory card
Function	activate, install, office, CAD, PS, software, systems, playing games, office, design, used, students, learning
After-sales service	customer services, refund, warranty, after-sales, invoice, repair, replacement
Commodity comparison	Which, comparison, better, cost performance

According to the cloud diagram of clustering words, the "gift" in the first kind of clustering question content comes from Fig. 6 (a). The "quality" in the second kind of question content can be obtained from Fig. 6 (b), (c), (d), (e) and (f). The "function" in the third kind of question content can be reflected from Fig. 6 (g) and (I). And "after-sales service" in the fourth kind of question content can be identified in Fig. 6 (h). As for the questions of "commodity comparison" in category 5, this paper also segment words and remove stop words, as shown in Fig. 7.

For the difference in each type of question content and mining information, this paper constructs the word2vec model and get the five kinds of main key input model. Finally, we get the 50 similar words from the word2vec model. And all similar words are back to the original text, which can correspond to actual classification of words. The results as shown in Table 3.

By exploring the questions of the Q&A system, we found that in the five categories of questions obtained by clustering, the number of quality related questions is the largest, which is followed by functions, gifts and commodity comparison. However, the number of questions related to after-sales service is the least. The average number of quality responses was the highest, which is followed by function and product comparisons and gifts. And the lowest number of quality responses is after-sales service. In terms of the mean of questions, the maximum quantity of words is the after-sales service while the least is the gift. In terms of the average number of answering words, the most word count is the function while the least count is the gift. The problems in Q&A system is the commodity information that consumers focus on during shopping. Given that the above results, we can conclude that the content that the consumer pays the most attention to or the content which causes the ambiguous meanings is quality. So these questions will also be more detailly described (the most number of average words). The average number of questions is low in the after-sales service where the consumer doesn't pay a lot of attention to, which is because most of the merchants' after-sales service quality and credibility is higher. When the warranty and after-sales guarantee can be provided, the rights and interests of consumers can be generally protected. Therefore, the number of questions in after-sales service is relatively small.

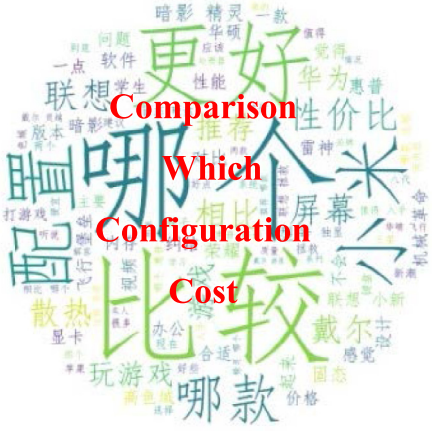


Fig. 7 Cloud diagram of commodity comparison.

4.2. Features of user generated content in online comments

By adopting unsupervised machine learning K-means algorithm to cluster of online comments, this paper found that the category of online comments is more than 5. So we set parameter n_clusters = range (5,20,1) to cluster the text data of online comments. According to the silhouette coefficients, the K value can be determined. Then the characteristics of the information content in e-commerce online comments are mined, as well as the information categories of online comments are determined.

In this paper, after several clustering experiments, redundant noise words were removed. The silhouette coefficient was calculated, and we obtain the best clustering effect when K = 8, as is shown in Fig. 8. Based on the clustering results, the final clustering results can be summarized into 6 categories, including appearance, logistics, gifts, quality, after-sales service and price. The clustering effect of online comments is shown in Table 4.

4.3. Comparison analysis in e-commerce Q&A and online comments

In this paper, information content features mined by online comment clustering and extracted by Q&A system are compared and analyzed, as is shown in Fig. 9. First, free gifts, quality and after-sales service are mentioned in the online review and Q&A system. We can find that consumers will still ask questions in the Q&A system even if the online comments have provided similar information. Second, online comments mainly focus on price, appearance, and logistics. This is because online comments are consumers' shopping and usage experience. Most of the price information will be displayed in the most prominent position, so consumers have no doubt in this regard. However, in the online comments, consumers usually have an intuitive feeling for commodity price and value after the purchase. These can be reflected in the comments information. For example, consumers generally don't have questions about logistics distribution for the laptops. However, consumers may make the corresponding evaluation of

Table 3 Different clustering categories in Q&A system.

Category of questions	Number of Q&A	Proportion	Mean of questions	Mean of question words	Mean of answers	Mean of answering words
Gifts	7718	9.7%	5.19	12.64	2.57	5.21
Quality	38,353	48.2%	25.81	23.06	4.01	7.64
Function	23,560	29.6%	15.85	17.33	3.32	9.68
After-sales service	3698	4.6%	2.49	11.59	2.21	6.68
Commodity comparison	6224	7.8%	4.19	16.45	3.36	8.36

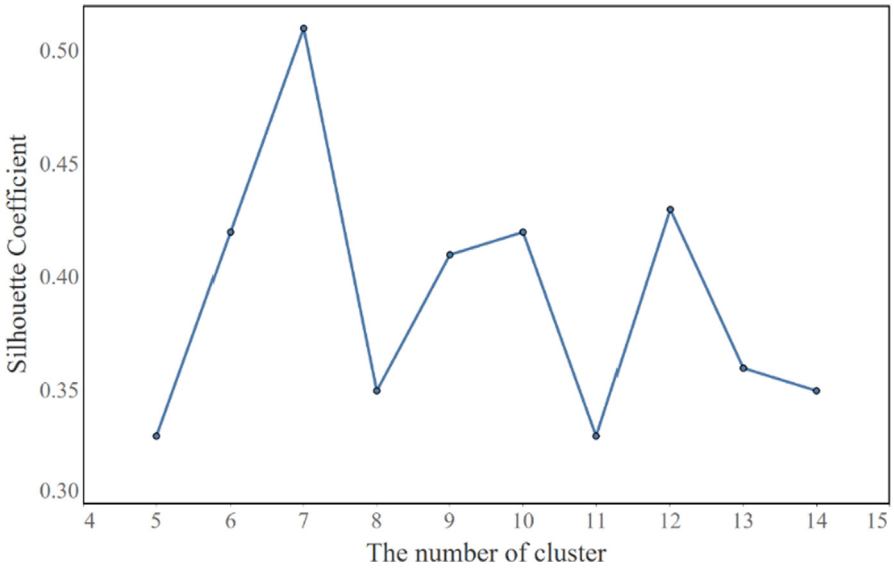


Fig. 8 Silhouette coefficients of different cluster numbers in online comments.

logistics speed and packaging quality in online comments. The appearance has been clearly shown in the product details, so consumers don't have too much doubt about the appearance of the computer. However, comments on the appearance can be reflected from online comments. Third, the unique content of the Q&A system is the question about the function and the comparison of goods. The comparison of goods means that consumers ask questions when it is difficult to make decision on similar goods. The function is to confirm whether the product can better match use requirements. It is known that the Q&A system and online comments are not only complementary to each other, but also provide verification for the information of online comments.

The question mining results of the Q&A system are compared with the online comments. This study found that the commodity comparation and function have no information provided in the online comments, but they are provided in Q&A system. Consumers can receive more information from the Q&A system and perceive the usefulness of Q&A system. Second, although the quality, free gifts and after-sales service provided in the Q&A system are also provided in the online comments, the surge in the number of comments and the inappropriate interference of merchants lead to the overload and distortion of online comments. Then consumers may lose truth in online comments. Therefore, on the one hand, consumers need to confirm this kind of information through the Q&A system of e-commerce. It can also be said that consumers believe that the content of the Q&A system of e-commerce is more reliable. On the other hand, consumers believe that it will take less time and effort to obtain information through the Q&A system, so it can be considered that the cognitive cost for consumers to obtain information through the Q&A system is lower.

Table 4 Feature word clustering in online comments.

Clustering categories	Main features
Appearance	Appearance, fashion, cool, handsome, beautiful, good-looking, thin
Logistics	Packaging, integrity, damage, logistics, delivery, express delivery, sealing, satisfaction
Gifts	Gifts, accessories, gifts, mouse, keyboard, computer bags, bags
Quality	Quality, screen, clarity, picture quality, running speed, fan, cooling, performance, noise, game effects, battery, battery life
After-sales service	After-sales service, return, replacement, after-sales, service
Price	Price, offer, discount, activity

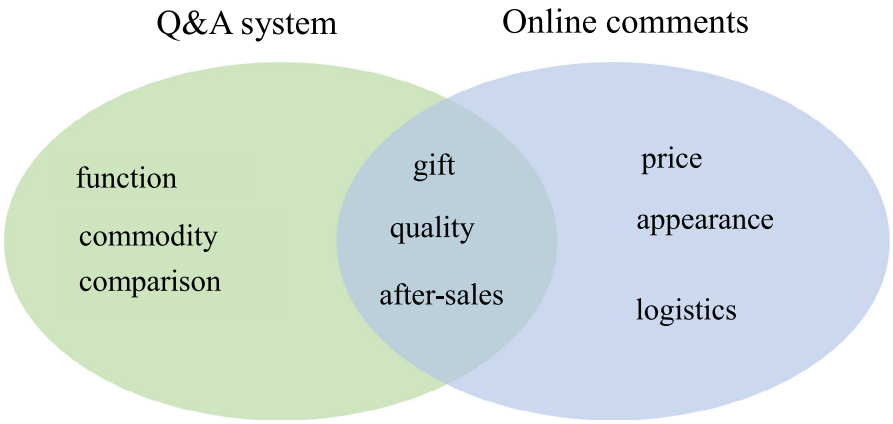


Fig. 9 Comparison of text content in Q&A system and online comments.

5. Conclusions

This paper analyzes the content characteristics of Q&A system and online comments through text classification and clustering, comparing the two kinds of user generated content to find similarities and differences. Then the attribute characteristics of Q&A system and online comments in information content are obtained. Furthermore, the research found that the Q&A system and online comments are not only complementary to each other, but also the Q&A system provide validation for online comments.

From the perspective of theoretical significance, this paper firstly classified and clustered user-generated content based on the LSTM and the improved K-means algorithm. Secondly, the clustering experiment in this paper is not only obtain several categories of text data, but also conduct the second clustering for each category based on the results of the first level of clustering. The improved K-means clustering algorithm revealed the text topic information in different depths, and realized the multi-level clustering from coarse granularity to fine granularity. Thirdly, the attributes of different user-generated content channels are summarized through the comparative analysis of the content characteristics of the Q&A system and online comments.

From the perspective of practical significance, the conclusions have important practical significance for e-commerce platforms, merchants and consumers. For the e-commerce platform, it is helpful to improve and perfect the Q&A system, and integrate the questions and answers of the Q&A system to find the most concerned questions of consumers. Then the more comprehensive and objective answers can be formed, which can save the time of users to read item by item and improve shopping experience. Second, it contributes to take measures to encourage consumers to participate in Q&A system platform according to the content characteristics. The more comprehensive questions and answers can be formed, then consumers can have a more comprehensive understanding of products. Third, it is helpful for merchants to identify the real needs of consumers, and find out what consumers are most concerned about. Then products can be improved in a targeted way to release accurate and attractive product information description, develop personalized marketing strategies, and finally improve consumers' shopping experience and advantages in the competition of similar products. From the perspective of consumers, online shopping will cost a lot of time to find the commodity information and comparison of similar goods, getting the information they need by browsing detailed descriptions, online reviews, and asking merchants in the past. For this reason, Q&A system can help consumers quickly find the questions and answers they need from the large amount of information, and then make the quick purchase decision after confirming the confused information.

There are still some deficiencies that need to be improved in the study. First, the online comments could be mined detailly. In the future, we can use more stop words list and better word segmentation when preprocessing the online comments of e-commerce. In the process of clustering, the algorithm with higher accuracy can be selected to make the information more comprehensive. Second, the word of mouth, operation mode and price in different e-commerce platform are different, which may lead to the deviation of the text content of Q&A system. In the future, text data of multiple e-commerce platforms also can be used for analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported in part by the Public Projects of Zhejiang Province under Grant (LGN20G010002, LGF19G010002, LGF19G010003), in part by Science and Technology Project of Zhejiang Province (2021C03143, 2021C02039).

References

[1] J. Hamari, M. Sjöklint, A. Ukkonen, The sharing economy: Why people participate in collaborative consumption, J. Assoc. Inform. Sci. Technol. 67 (9) (2015) 2047–2059.

[2] W. Kathan, K. Matzler, V. Veider, The sharing economy: Your business model's friend or foe?, Bus Horiz. 59 (6) (2016) 663–672.
[3] X. Fang, J. Wang, D. Seng, B. Li, C. Lai, X. Chen, Recommendation algorithm combining ratings and comments, Alexandria Eng. J. 60 (6) (2021) 5009–5018.
[4] M. Chen, G. Chao, X. Ding, Impact of online comments on purchase intention of college student consumers under online shopping, Asian Agric. Res. 8 (12) (2016) 29–34.
[5] E.C. Webb, Understanding the Use of Online Reviews and Recommendations in Consumer Judgment and Decision-Making, Adv. Consum. Res. 45 (2017) 302–306.
[6] P.F. Christopher, A.Z. Robert, The influence of information overload on the development of trust and purchase intention based on online product reviews in a mobile vs. web environment: an empirical investigation, Electronic Markets 27 (3) (2017) 211–224.
[7] K.C. Soylemez, Impact of individual and brand level factors in generation of different user-generated content, J. Consum. Market. 38 (4) (2021) 457–466.
[8] S. Tirunillai, G.J. Tellis, Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance, Soc. Sci. Electronic Publ. 31 (2) (2012) 198–215.
[9] C. Shi, P. Hu, W. Fan, L. Qiu, How learning effects influence knowledge contribution in online Q&A community? A social cognitive perspective, Decis. Support Syst. 149 (2021) 113610.
[10] E. Choi, C. Shah, User motivations for asking questions in online Q&A services, J. Assoc. Inform. Sci. Technol. 67 (5) (2016) 1182–1197.
[11] J. Yu, Z. Jiang, H.C. Chan, The Influence of Socio technological Mechanisms on Individual Motivation toward Knowledge Contribution in Problem-Solving Virtual Communities, IEEE Trans. Prof. Commun. 54 (2) (2011) 152–167.
[12] C. Fang, J. Zhang, Users Continued Participation Behavior in Social Q&A Communities: A Motivation Perspective, Comput. Hum. Behav. 92 (2019) 87–109.
[13] Y. Meng, H. Wang, L. Zheng, Impact of online word-of-mouth on sales: the moderating role of product review quality, New Rev. Hypermedia Multimedia 11 (2018) 1–27.
[14] C. Changchit, T. Klaus, Determinants and Impact of Online Reviews on Product Satisfaction, J. Internet Commer. 19 (1) (2020) 82–102.
[15] F. Hu, The relationship analysis between online reviews and online shopping based on B2C platform technology, Cluster Comput. 22 (S2) (2019) 3365–3373.
[16] W. Zhou, W. Duan, Do Professional Reviews Affect Online User Choices Through User Reviews? An Empirical Study, J. Manage. Inform. Syst. 33 (1) (2016) 202–228.
[17] J.R. Mosteller, C. Mathwick, Online Reviewer Engagement: A Typology Based on Reviewer Motivation, J. Serv. Res. 20 (2) (2017) 204–218.
[18] Y. Chen, J. Xie, Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix, Manage. Sci. 54 (3) (2008) 477–491.
[19] S. Mitra, M. Jenamani, Helpfulness of Online Consumer Reviews: A Multi-Perspective Approach, Inf. Process. Manage. 58 (3) (2021) 102538.
[20] Y. Heng, Z. Gao, Y. Jiang, X. Chen, Exploring hidden factors behind online food shopping from Amazon reviews: A topic mining approach, J. Retailing Consum. Serv. 42 (2018) 161–168.
[21] A. Ghose, P.G. Ipeirotis, Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics, IEEE Trans. Knowl. Data Eng. 23 (10) (2011) 1498–1512.
[22] G. Qiu, B. Liu, J. Bu, C. Chen, Opinion Word Expansion and Target Extraction through Double Propagation, Comput. Linguistics. 37 (1) (2011) 9–27.
[23] F. Lazhar, Mining hidden opinions from objective sentences, Int. J. Data Min., Modell. Manage. 10 (2) (2018) 113–126.
[24] M. Uriarte, R.L. Chazdon, Incorporating natural regeneration in forest landscape restoration in tropical regions: synthesis and key research gaps, Biotropica 48 (6) (2016) 915–924.
[25] Z. Liu, Q. Shen, J. Ma, Z. Dong, Research on comment target extracting in Chinese online shopping platform, Int. J. Crowd Sci. 2 (3) (2018) 247–258.
[26] W. Shu, K. Cai, N.N. Xiong, Research on Strong Agile Response Task Scheduling Optimization Enhancement with Optimal Resource Usage in Green Cloud Computing, Future Gener. Comput. Syst. 124 (2021) 12–20.
[27] B.S. Kumar, V. Ravi, A survey of the applications of text mining in financial domain, Knowl.-Based Syst. 114 (2016) 128–147.
[28] B.H. Ye, J.M. Luo, H.Q. Vu, Spatial and temporal analysis of accommodation preference based on online reviews, J. Destination Market. Manage. 9 (2018) 288–299.
[29] A.K. Uysal, S. Gunal, The impact of preprocessing on text classification, Inf. Process. Manage. 50 (1) (2014) 104–112.
[30] C.X. Jin, H.Y. Zhou, Q.C. Bai, Short Text Clustering Algorithm with Feature Keyword Expansion, Adv. Mater. Res. 532-533 (2012) 1716–1720.
[31] S. Farzi, H. Faili, Improving Statistical Machine Translation using Syntax-based Learning-to-Rank System, Digital Scholars. Human. 32 (1) (2017) 80–100.
[32] X. Luo, Efficient English text classification using selected Machine Learning Techniques, Alexandria Eng. J. 60 (3) (2021) 3401–3409.
[33] K. Ghany, H. Zawbaa, H. Sabri, COVID-19 prediction using LSTM Algorithm: GCC Case Study, Informatics in Medicine Unlocked, 23 (2021) 100566.
[34] Y. Jang, I.B. Jeong, Y.K. Cho, Y. Ahn, Predicting Business Failure of Construction Contractors Using Long Short-Term Memory Recurrent Neural Network, J. Construct. Eng. Manage. 145 (11) (2019) 1–9.
[35] S. Goudarzi, N. Kama, M.H. Anisi, S. Zeadally, S. Mumtaz, Data collection using unmanned aerial vehicles for Internet of Things platforms, Comput. Electr. Eng. 75 (2019) 1–15.
[36] M. Ali, S. Qaisar, M. Naeem, S. Mumtaz, Energy Efficient Resource Allocation in D2D-Assisted Heterogeneous Networks with Relays, IEEE Access 4 (2016) 4902–4911.
[37] B. Ji, X. Zhang, S. Mumtaz, C. Han, C. Li, H. Wen, D. Wang, Survey on the Internet of Vehicles: Network Architectures and Applications, IEEE Commun. Stand. Magaz. 4 (1) (2020) 34–41.